DOCUMENT RESUME

ED 175 265 FL 010 534

AUTHOR Larsen-Freeman, Diane
TITLE Assessing Proficiency Through an Analysis of L2 Learner Speech.
PUB DATE 78
NOTE 16p.; Paper presented at the Second Language Acquisition Forum (Los Angeles, California, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Communicative Competence (Languages); *Error Analysis (Language); Error Patterns; *Language Patterns; *Language Proficiency; Language Research; Language Skills; Language Tests; Measurement; Predictive Measurement; *Second Language Learning; Speech Communication; *Speech Skills; Syntax
IDENTIFIERS Interlanguage

ABSTRACT

The need for an index of development by which second language proficiency could be expediently and reliably gauged has been acknowledged by both second language (L2) teachers and researchers. In two previous L2 studies, the search for an index of development centered on an examination of learner written performance. In an attempt to construct an index of development that is valid regardless of performance modality, the study reported here was initiated for the purpose of analyzing features of learner speech. Twenty-five learners of English as a Second Language (ESL), comprising different L1 backgrounds and five levels of proficiency, were asked to perform four speech tasks. Their responses were recorded and transcribed. The data analysis showed that the average length of error-free T-units appeared to discriminate among learners of differing proficiencies. It is hoped that findings from both phases will eventually provide researchers and teachers with an effective tool by which to assess the progress of a second language learner's development. (Author/AMH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ASSESSING PROFICIENCY THROUGH AN ANALYSIS OF L2 LEARNER SPEECH[1]

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Diane Larsen-Freeman
School for International Training
The Experiment in International Living

ED175265

INTRODUCTION

This paper is the third in a series of reports on the progress we are making in our attempt to establish a second language acquisition index of development. An index of development, a term I have borrowed from Hakuta (1975), is a developmental yardstick against which second language proficiency can be measured. Such an index would allow us to give numerical values to different points of a learner's interlanguage--points along the L2 developmental continuum--much as the MLU or mean length of utterance measure does in L1 acquisition. Ideally, these numerical correlates of the developmental process would increase uniformly and linearly as a learner moved towards full acquisition of the target language.

The reason for our efforts to construct such an index is our recognition of the difficulty in comparing research findings from studies where the subjects' proficiency is characterized by the imprecise terms "beginning", "intermediate" and "advanced". Such relative terms do not allow meaningful comparisons to be made across subject population. What is beginning to me might not be to others. What we have been attempting to do in a series of studies is to find an objective measure by which to assess and report English as a second language proficiency. Our hope was that if we could find an ESL index that was not based upon the structure of English itself, we might be able to employ the same developmental index for other target languages.

FL010624

---

[1]Many thanks to Cheryl Andersen Butoyi, Hossein Farhady and Virginia Strom for their help in the research project reported on in this paper.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Diane Larsen-Freeman

© Diane Larsen-Freeman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The two earlier reports we issued were based on analyses of ESL learners' writing. From these analyses we learned that the measures which best satisfied our criteria of objectivity, reliability, validity and universality were measures based on Kellogg Hunt's T-unit performance variable. A T-unit is "one main clause plus whatever subordinate clauses, phrases and words happen to be attached to or embedded within it." (Street: 1971:13) So the sentence, "I've got a dog and his name is Blue," contains two T-units, whereas the sentence, "I've got a dog named Blue," only contains one. T-unit length worked better than sentence length as a measure of syntactic maturity, Hunt found, because a sign of writing immaturity, the so-called "run-on" sentence, can be divided into two or more T-units. The three performance variables which were most suitable contenders for a developmental index were: the total number of error-free T-units, the % of error-free T-units and the average length of error-free T-units as written by ESL students in English compositions. Thus we discovered that while average T-unit length was a reliable index of L1 syntactic maturity, introducing the notion of "error-free T-unit" made an index of L2 syntactic proficiency more powerful.

The next study we conducted, and the one reported on here, was initiated to determine if these same measures could be applied to ESL learners' speech in order to discriminate among ESL proficiency levels. Hunt (1965) and other first language acquisition researchers (O'Donnell, Griffin and Norris, 1967; Loban, 1976) found T-unit length to be a highly satisfactory index of oral as well as written complexity. In second language acquisition, Thornhill (1969) used two different oral tasks (an interview and oral responses to a series of pictures) to study the development over a nine-week period of syntactical fluency of four adult Spanish speakers

learning ESL. Thornhill computed the average number of words per T-unit and noted that there was an overall increase in T-unit length by subjects during this nine-week period. However, the increases were irregular for all four subjects--i.e., for all subjects the lowest number of words was recorded during the first interview, but for none of the four was the highest number of words recorded during the last interview. (36) Despite the irregularity, Thornhill concluded "the mean length of T-units (is) a usable measure of development towards maturity in second language production." (1969:37).

None of these researchers, however, applied the notion of "error-free T-unit," the construct we had found to work well. It was our hope that such a construct would be the answer to our search for an L2 acquisition index of development.

It seemed to us that T-unit length itself made some sense since elaboration of T-units calls for a learner to increasingly control the processes of language production--addition, deletion, substitution and permutation. By deleting redundant NP's in compound sentences, embedding less relevant information, relating two or more sentences using logical connectors and many other operations, a learner could increase T-unit length.

Yet, despite our enthusiasm for T-unit length, we agreed with Gaies who argued that mere length of T-units might not be enough for ascertaining proficiency. Gaies observed that "some account must be taken of structural errors which occur in L2 learners' writing, for they, too, are an indication of incomplete syntactic control, just as much perhaps as is oversimplified (i.e., short) sentence structure." (1976:7) Thus, syntactic elaboration, combined somehow with the elimination of error production, seemed intuitively and empirically to have a great deal of appeal as a basis for an index of

development. We decided to conduct a study using speech samples to probe this construct further.

SUBJECTS

For this study 25 subjects were used: five each from each of five different proficiency levels. The five different proficiency levels were established by the subjects' performance on the UCLA English as a Second Language Placement Examination (ESLPE)--a statistically well-established instrument, having a high internal consistency reliability of .97. Those subjects who placed in Group 1 had scored the lowest on the exam thus indicating they required a great deal of ESL instruction before being permitted to freely join the academic mainstream; those who placed in Group 5 were those subjects who scored sufficiently highly on the placement exam to have all ESL requirements waived.

These subjects, both graduate and undergraduate students at UCLA, represented 11 different L1 backgrounds as follows: 7 Chinese speakers, 5 Farsi speakers, 3 Spanish speakers, 2 French speakers, 2 Russian speakers, 1 speaker each from Arabic, Greek, Indonesian, Japanese, Korean and Tagalog native languages.

TASKS

Each subject was given four oral tasks to perform. They were administered individually. They were as follows:

I. Interview

Each subject was asked his name, native country, how long he had been in the United States, what his major was, how long he had been studying

English, and what his future plans were. These were the basic questions asked of all subjects. Subjects were also asked follow-up questions embroidering on their responses to these basic questions.

II. Picture Composition Task

Using four picture series from Donn Byrne's book, Progressive Picture Compositions, each subject was asked to relate the chain of events in each of the four picture series. Prior to each subject beginning the task, the experimenter modeled a different picture series.

III. Conversation

Using anything the experimenter picked up of interest from the student's comments during the interview, the experimenter initiated a conversation with each student. The other topics brought in and discussed at this time were:

- the subject's impressions of Los Angeles
- any travel the subject had done while in the United States
- an elaboration of the composition topic each subject had written about earlier which was "Should a person live in a large city or a small town?"

IV. Story Retelling

The experimenter read a story to each subject once (a slightly modified version of "The Brahman and His Two Wives"--from International Folktales I by Binner) and each subject was asked to retell the story using his own words.

These four tasks were performed by each subject with one of two experimenters doing the administration. Each subject's responses were tape-recorded and later transcribed for analysis. The analysis itself was performed using

the O'Donnell, Griffin and Norris (1967) procedure for analyzing speech samples.

RESULTS

I. Interview

For the first task, the interview, we conducted an error analysis of the transcripts to see if the total number of errors or different error types were useful in discriminating among proficiency levels. We weren't optimistic about these since we had used this same tact when we analyzed writing samples and found little of significance, but we didn't want to overlook any possibility.

We devised a classification schema consisting of 29 categories of errors. Each error we found in the transcripts was placed into one of these categories. We discovered that the different groups of subjects essentially committed the same types of errors. Errors involving prepositions, articles and verb tense were most prevalent. The other categories were not as consistently ranked after these, but errors in the top ten error types for each level did account for approximately 80% of the total number of errors. The other 20% of the errors were accounted for by the remaining 19 error types.

Curiously enough, as far as the interview was concerned, even the sheer number of errors committed by the different groups did not distinguish among proficiency levels. While level 1 committed 94 errors and level 2, 42, level 3 committed 67, level 4, 78 and level 5, 76. Thus, there was no reduction in the number of errors made as proficiency level increased. One explanation for this surprising finding is that although we asked the same questions of each subject, and each interview lasted

approximately the same length of time, the more advanced subjects elaborated each answer more. In fact, they produced more utterances, thus creating more occasions for errors to be made.

II. Picture Composition

For Task II, the picture composition task, we also conducted an error analysis. This time, perhaps due to this being a more structured task, the numbe· of errors did decline as the proficiency level increased. Level 1 committed 207 errors, level 2, 168, level 3, 134, level 4, 130 and level 5, 88.

It is also interesting to note that while errors involving prepositions and articles were still among the most prevalent error types, the error types of verb tense, subject-verb agreement and lexical choice surpassed them in frequency of occurrence.

Apparently when given a picture composition to relate, a more structured task than an interview, avoidance is more difficult and error types of a different kind surface. Of course, it is also possible that it is easier to spot errors of a certain type when the context is clearly defined.

Let us now turn to the quantitative analyses we conducted. The first we'll consider is the one we conducted of the transcript from the picture composition task. We first correlated six performance variables with our subjects' scores on the ESLPE. The six performance variables were: number of words, number of T-units, number of error-free T-units, number of words per T-unit and the average number of words per error-free T-unit. (See Table 1.)

Table 1 Correlations between ESLPE scores and performance variables

	Picture A	Picture B	Picture C	Picture D	Combined Pictures
# of words	.0629 S=.383	.2021 S=.166	-.0956 S=.325	.2695 S=.096	.1652 S=.215
# of T-units	-.2845 S=.084	.1195 S=.285	-.1992 S=.170	.0266 S=.450	-.0885 S=.337
# of error-free T-units	.2426 S=.121	.5673** S=.002	.2411 S=.123	.4108* S=.021	.4811** S=.007
# of words in error-free T-units	.3933* S=.026	.6832** S=.001	.4013* S=.023	.5031** S=.005	.5859** S=.001
Average # of words/T-unit	.5209** S=.004	.0914 S=.332	.0402 S=.424	.3288 S=.054	.3570* S=.040
Average # of words/error-free T-unit	.4093* S=.021	.4988** S=.006	.4213* S=.018	.3302 S=.053	.5725** S=.001

*$p < .05$

**$p < .01$

As can be seen in Table 1, we obtained significant correlations at .01 in 3 of our performance variables: number of error-free T-units, number of words in error-free T-units and the average number of words per error-free T-unit. The most promising measure was the average number of words per error-free T-unit (the bottom line) since the number of error-free T-units and the number of words in error-free T-units would be affected by length of output. This was unacceptable since we wanted some measure which could be applied to

most any output independent of length or task.

Focusing on average length of error-free T-unit, we next conducted an analysis of variance. (See Table 2.)

Table 2 ANOVA - Average number of words/error-free T-unit (Combined Pictures)

5 groups			3 groups			2 groups		
n	mean	SD	n	mean	SD	n	mean	SD
5	5.76	2.20	8	6.69	.97	14	6.62	1.80
5	7.07	1.11	8	7.10	1.45	11	8.34	2.02
5	6.94	1.79	8	8.97	2.03			
5	8.18	2.48						
5	8.93	1.57						
F ratio = 2.07			F ratio = 4.93			F ratio = 5.07		
F prob. = .123			F prob. = .0176*			F prob. = .034*		

Post hoc analysis:	t obs
G1 vs. G2 NS	.93
G2 vs. G3 SIG	4.32
G1 vs. G3 SIG	5.11
t crit = 2.61	

*$p < .05$

We learned that an n of 5 per group was too small to produce a significant analysis of variance. Curious as to what size n one would need to obtain significant results, we next divided our 25 subjects into two groups--those above the mean ESLPE score and those below it. When we reduced the groups to two, we did indeed discover that significant differences resulted. Heartened by this finding, we next conducted an ANOVA with three groups, each having an n of 8. A significant F ratio was obtained. Next we applied

Cheffee's method of post-hoc analysis to see whether each group was significantly different from the others. This analysis revealed that groups 2 and 3 and groups 1 and 3 were significantly different from each other.

III. Conversation

We had discovered in our earlier study of ESL students' compositions that sheer length of output was distinctive. In general, more advanced learners tended to produce longer compositions than less proficient learners. However, at a certain developmental point, there was a leveling off and the most proficient learners did not necessarily produce the longest compositions. We applied a measure of length to the conversation task to see if this characterization would hold true for the oral modality as well. (See Table 3.)

Table 3 ANOVA - Total number of words

5 groups			3 groups			2 groups		
n	mean	SD	n	mean	SD	n	mean	SD
5	293.79	106.93	8	252.37	112.19	14	285.50	151.06
5	234.20	121.02	8	370.12	172.70	11	381.81	162.77
5	327.20	205.21	8	376.25	180.73			
5	424.59	152.72						
5	359.59	193.31						
F ratio = .981			F ratio = 1.558			F ratio = 2.340		
F prob. = .44			F prob. = .23			F prob. = .13		

Although we didn't achieve significance with any of our ANOVA's, it is interesting to note that the general pattern regarding length is the same as it was for the composition, i.e., the length of speaking increases gradually for the first four groups (although there is some backsliding

between the first and second groups) and then length drops off with the fifth group. We speculated with regards to writing that the top group perhaps didn't feel the need to demonstrate their competence by sheer volume of output, and instead confined their energies more to refining the final product. Perhaps this is the case here as well.

IV. Story Retelling

We conducted basically the same quantitative analysis that we did for the picture composition task for the story retelling task. The same three performance variables once again, in general, correlated significantly with the subjects' scores on the ESLPE. (See Table 4.)

Table 4 Correlation between ESLPE scores and performance variables

# of words	.3131 S=.064	# of T-units	.1847 S=.188
# of error-free T-units	.5031 S=.005**	# of words in error-free T-units	.6260 S=.001**
Average # of words/T-unit	.2710 S=.095	Average # of words/error-free T-unit	.5057 S=.005**

**p < .01

Taking the average number of words per error-free T-unit as the measure which best satisfies the criteria which were mentioned earlier, we again performed an analysis of variance.

Although none of the analyses were significant, the analysis with two groups was not too far off. Furthermore, we can see from each analysis that the mean length of error-free T-unit did increase in length with each

increase in proficiency level. (See Table 5.)

Table 5 ANOVA - Average number of words/error-free T-unit

5 groups			3 groups			2 groups		
n	mean	SD	n	mean	SD	n	mean	SD
5	7.39	2.06	8	7.83	1.68	14	7.77	1.67
5	7.56	1.84	8	8.28	0.84	11	9.29	2.98
5	8.21	1.04	8	9.73	3.41			
5	8.44	1.03						
5	10.61	4.10						
F ratio = 1.552			F ratio = 1.565			F ratio = 2.608		
F prob. = .2257			F prob. = .2326			F prob. = .1200		

Since the ANOVA was approaching significance with two groups, one could reasonably assume that significance could be reached given a higher n so at least this measure, the average length of error-free T-units, seems suitable for distinguishing among proficiency levels for groups of learners.

CONCLUSION

From all available evidence, it appears that the average length of T-units is a highly satisfactory index of syntactic maturity for L1 acquisition. Yet we have learned from our own research that a more powerful measure of L2 proficiency is the length of error-free T-units. L2 learners are conceivably capable of producing T-units of some length before they can be said to have truly acquired the L2. By assessing a learner's proficiency through a quantitative analysis of a learner's error-free T-units, a learner's score cannot be artificially inflated by his simply

stringing words together.

We have one more series of task results to analyze and report on, but we are becoming increasingly comfortable with the average length of error-free T-units as being an L2 criterion measure analogous to the MLU in L1 acquisition. It seems to work for two modalities for distinguishing among L2 proficiency levels with ESL students at the college level. If we do indeed conclude that this measure is the answer to our search, we would also want to determine if it works equally well for different populations of L2 learners and with learners studying different target languages. Nonetheless, at this point in time, we are somewhat encouraged that we may have found a measure which works well at least for groups of ESL learners at the college level.

REFERENCES

Gaies, S. Sentence-combining: a technique for assessing proficiency in a second language. Paper presented at the Conference on Perspective on Language, University of Louisville, Louisville, Kentucky, 1976.

Hakuta, K. Becoming bilingual at age five: the story of Uguisu. Thesis, Harvard University, 1975.

Hunt, K. Grammatical structures written at three grade levels, NCTE Research Report No. 3, Champaign, Ill.: National Council of Teachers of English, 1965.

Larsen-Freeman, D. Evidence of the need for a second language acquisition index of development. In Ritchie, W. (Ed.), Second Language Acquisition Research. New York: Academic Press, 1976.

Larsen-Freeman, D. An ESL index of development. TESOL Quarterly, 1978, 12, 4.

Larsen-Freeman, D. and Strom, V. The construction of a second language acquisition index of development. Language Learning, 1977, 27, 1, 123-134.

Loban, W. Language development: kindergarten through grade twelve. NCTE Research Report No. 18, Urbana, Ill.: National Council of Teachers of English, 1976.

O'Donnell, R., Griffin, W. and Norris, R. Syntax of kindergarten and elementary school children: a transformational analysis. NCTE Research Report No. 8, Champaign, Ill.: National Council of Teachers of English, 1967.

Street, J. Readability of UCLA materials used by foreign students. M.A. - TESL thesis, UCLA, 1971.

Thornhill, D. A quantitative analysis of the development of syntactical fluency of four young adult Spanish speakers learning English. Ph.D. dissertation, Florida State University, 1969.